

D6.2 Data Management Plan

Deliverable ID:	D6.2
Project acronym:	E-CONTRAIL
Grant:	101114795
Call:	HORIZON-SESAR-2022-DES-ER-01
Topic:	HORIZON-SESAR-2022-DES-ER-01-WA1-6
Consortium coordinator:	Universidad Carlos III de Madrid
Edition date:	06 October 2023
Edition:	01.00
Status :	Draft
Classification:	PU

Abstract

The objective of this Deliverable is to describe the Data Management Plan (DMP) of E-CONTRAIL Project. E-CONTRAIL is included in the framework of the Digital European Sky research and innovation programme of the SESAR 3 Joint Undertaking.

The Deliverable includes a data summary, including the purpose of data to fulfil E-CONTRAIL objectives, a description of types and formats of input and output data, the origin, accessibility and approximate size of the data, the re-use of data and their utility for different target audiences. The deliverable also states a plan for continuous update of the document (with versions and content to be updated). E-CONTRAIL will follow FAIR data principles, and, thus, the deliverable elaborates on a set of guiding principles to make data Findable, Accessible, Interoperable, and Reusable. Allocation of resources to comply with FAIR principles and the data management strategy of E-CONTRAIL is provided within the document. Finally, an analysis on data security, including aspects related to data sharing and data storage is covered in the Deliverable.

Authoring & approval

Author(s) of the document

Organisation name		Date
BIRA	WP4 leader	21/08/2023
UC3M	Project Coordinator	29/08/2023

Reviewed by

Organisation name		Date
UC3M	WP1 contributor	08/09/2023
RMI	WP2 leader	30/08/2023
KTH	WP3 contributor	15/09/2023
BIRA	WP4 contributor	21/08/2023

Approved for submission to the SESAR 3 JU by¹

Organisation name		Date
UC3M	Project Coordinator	20/09/2023
RMI	WP2 leader	20/09/2023
KTH	WP3 leader	20/09/2023
BIRA	WP4 leader	20/09/2023

Rejected by²

Organisation name	Date
-------------------	------

Document history

Edition	Date	Status	Company author	Justification
00.00.01	11/08/2023	Initial Draft	UC3M	New document
00.00.02	21/08/2023	Contribution to Draft	BIRA	Sent for internal review
00.00.02	29/08/2023	Revision Draft	UC3M	Check of the entire draft
00.00.03	30/08/2023	Revision of draft	RMI	Revision of draft (GEO & ECMWF data)
00.00.04	15/09/2023	Revision of draft	KTH	Revision of draft

00.00.05	20/09/2023	Reviewed Draft	BIRA	Internally reviewed
00.01.0	06/10/2023	Submission	UC3M	Sent for internal review after SJU comments

Copyright statement © (2023) – (UC3M, BIRA, RMI, KTH). All rights reserved. Licensed to SESAR 3 Joint Undertaking under conditions.

E-CONTRAIL

ARTIFICIAL NEURAL NETWORKS FOR THE PREDICTION OF CONTRAILS AND AVIATION INDUCED CLOUDINESS

E-CONTRAIL

This document is part of a project that has received funding from the SESAR 3 Joint Undertaking under grant agreement No 101114795 under European Union's Horizon Europe research and innovation programme.



We provide a high-level summary of the project E-CONTRAIL:

Contrails and aviation-induced cloudiness effects on climate change show large uncertainties since they are subject to meteorological, regional, and seasonal variations. Indeed, under some specific circumstances, aircraft can generate anthropogenic cirrus with cooling. Thus, the need for research into contrails and aviation-induced cloudiness and its associated uncertainties to be considered in aviation climate mitigation actions becomes unquestionable.

We will blend cutting-edge AI techniques (deep learning) and climate science with application to the aviation domain, aiming at closing (at least partially) the existing gap in terms of understanding aviation-induced climate impact.

The overall purpose of E-CONTRAIL project is to develop artificial neural networks (leveraging remote sensing detection methods) for the prediction of the climate impact derived from contrails and aviation-induced cloudiness, contributing, thus, to a better understanding of the non-CO2 impact of aviation on global warming and reducing their associated uncertainties as essential steps towards green aviation.

Specifically, the objectives of E-CONTRAIL are:

- O-1 to develop remote sensing algorithms for the detection of contrails and aviation-induced cloudiness.
- O-2 to quantify the radiative forcing of ice clouds based on remote sensing and radiative transfer methods.
- O-3 to use of deep learning architectures to generate AI models capable of predicting the radiative forcing of contrails based on data- archive numerical weather forecasts and historical traffic.
- O-4 to assess the climate impact and develop a visualization tool in a dashboard.

Table of contents

1	<i>Data summary</i>	8
2	<i>FAIR data</i>	16
3	<i>Allocation of resources</i>	22
4	<i>Data security</i>	23
5	<i>Ethics</i>	25
6	<i>Other issues</i>	26

List of figures

Figure 1: Block Diagram of E-CONTRAIL project	8
Figure 2: Input/output data types	9
Figure 3: Open Aire portal corresponding to a Project coordinated by UC3M (ALARM) as an example of the one to be created for E-CONTRAIL	14
Figure 4: Unique and persisted identifier: Server (Servidor); Repository ID (Identificador del repository); Item ID (Identificador del ítem)	17
Figure 5: Type of archives	17
Figure 6: Dublin Core Metadata Terms	18
Figure 7: Creative Commons Attribution 4.0 International	21
Figure 8: GNU General Public License v3.0	21

List of tables

Table 1: list of acronyms	6
Table 2: E-CONTRAIL consortium acronyms	7
Table 3: E-CONTRAIL data input (description)	10
Table 4: E-CONTRAIL data output (description)	12

List of acronyms

Non-exhaustive list of acronyms used across the text.

Acronym	Description
AI	Artificial Intelligence
ATM	Air Traffic Management
CA	Consortium Agreement
COBRA	COvariance-Based Retrieval Algorithm
DML	Data Machine Learning
DMP	Data Management Plan
ECMWF	European Centre for Medium-Range Weather Forecasts
EU	European Union
EUMETSAT	European operational satellite agency for monitoring weather, climate and the environment from space
FAIR	Findable, Accessible, Interoperable and Reusable
FCI	Flexible Combined Imager
GA	Grant Agreement
GEO	Geostationary Earth Orbit
JU	Joint Undertaking
MSG	Meteosat Second Generation
MTG	Meteosat Third Generation
NetCDF	Network Common Data Form
NRT	Near-Real Time
NWP	Numerical Weather Prediction
ORD	Open Research Data
PMP	Project Management Plan
RGB	Red Green Blue
SESAR	Single European Sky ATM Research
SEVIRI	Spinning Enhanced Visible and InfraRed Imager
SJU	SESAR Joint Undertaking
WP	Work Package

Table 1: list of acronyms

E-CONTRAIL Consortium

Acronym	Description
BIRA	Royal Belgian Institute for Space Aeronomy
KTH	Royal Institute of Technology
RMI	Royal Meteorological Institute of Belgium
UC3M	Universidad Carlos III of Madrid

Table 2: E-CONTRAIL consortium acronyms

1 Data summary

1.1 Purpose

The main goal of the DMP is to provide an analysis of the main elements of the data management policy that the E-CONTRAIL consortium will use to comply with the established goals.

Recall that the overall goal of E-CONTRAIL is to develop artificial neural networks (leveraging remote sensing detection methods) for the prediction of the climate impact derived from contrails and aviation-induced cloudiness (AIC). Figure 1 presents a methodological block diagram sketching the different components of the project.

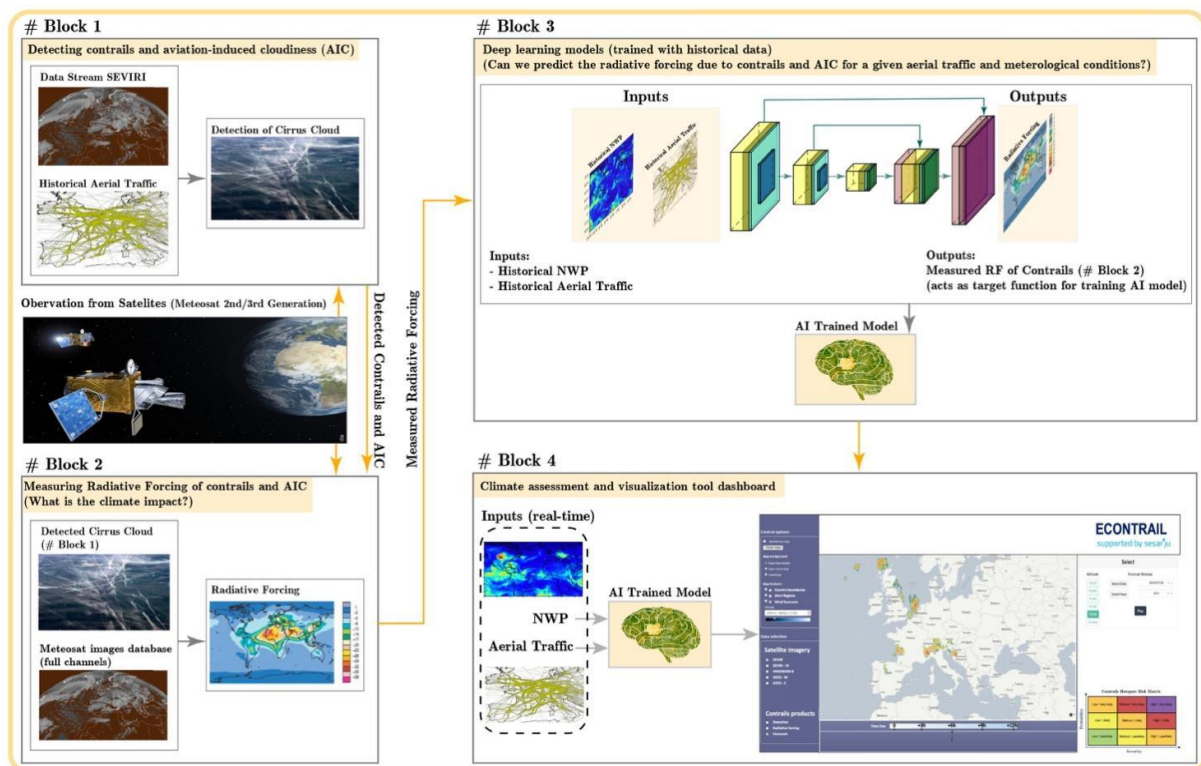


Figure 1: Block Diagram of E-CONTRAIL project

It comprises the (near-real-time – NRT) continuous detection of contrails and AIC from satellites (#Block 1 of Figure 1), with the objective to estimate the effective radiative forcing (ERF) related to these detections (#Block 2). NRT observations (with delay of delivery from 10 to 30 minutes) from polar Geostationary Earth Orbit (GEO) satellites, is used to feed models algorithms to discriminate natural cloudiness from induced cloudiness are needed (#Block 3). The resulting estimation are passed over as target functions to the AI module, which exploits deep learning to find the spatial and temporal relations between the inputs (in this case, numerical weather forecasts and traffic demand) and the output, which is the target function (in this case, the estimated radiative forcing). A visualisation tool will be assessed to illustrate related climate impact (#Block 4).

In light of this, it becomes apparent that the availability of data is essential to fulfil with all the activities of the project (data access, data storage/process, data output). Thus, they are essential to successfully conduct the research activities of the whole project.

1.2 Types of formats

Figure 2 sketches the input/output data scheme of E-CONTRAIL project. In summary, we use information from GEO satellite observations, historical aerial traffic data and Numerical Weather Predictions (NWP) forecasts. A Table which summarises all these inputs is presented in Section 1.2.1, and a description of these resulting outputs is given in Section 1.2.2.

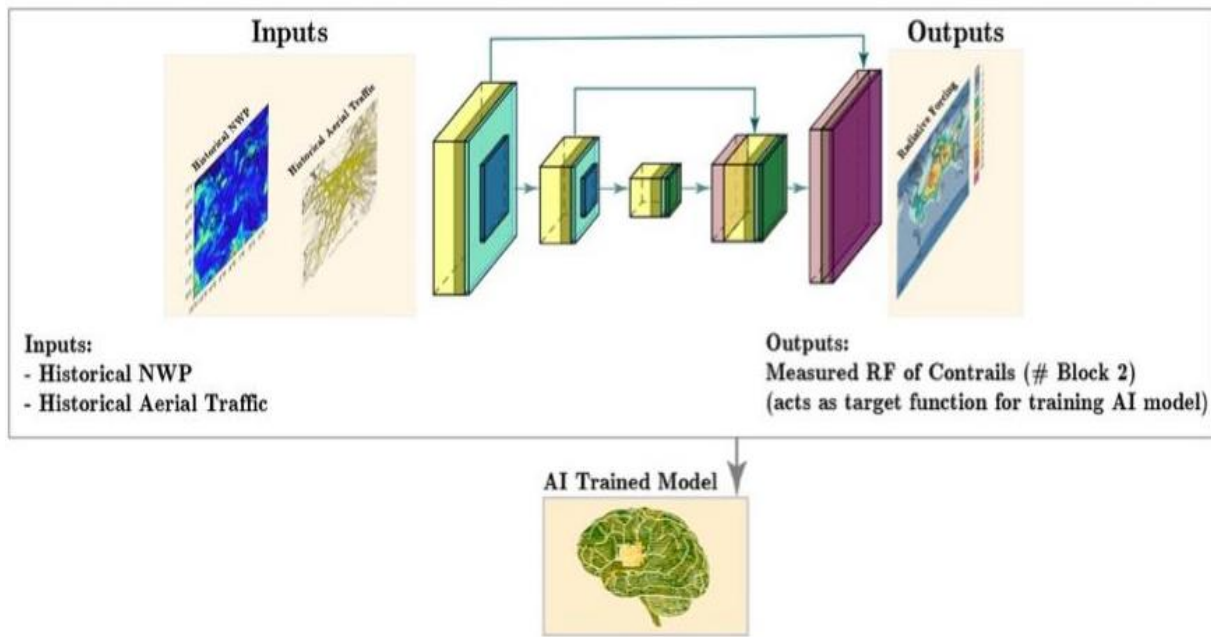


Figure 2: Input/output data types

1.2.1 Input data

As input data, E-CONTRAIL will¹ make use of the data sources to forecasts contrails, as described in Section 1.3. The E-CONTRAILS AI model will centralise a prototype system capable of integrating different data sources (GEO satellite detection of contrails and AIC, historical air traffic data and historical NWP forecasts), to develop nowcasting/forecasting of contrails.

¹ Detailed description of the data, including its origin, size, access rights (if any), utility and metadata (if any) will be provided as appendixes to this deliverable in the periodical updates (see Section 1.2)

Category	Data	Description
Particles in suspension (ice crystal)	Satellite observations	<ul style="list-style-type: none"> Data observations from MSG (SEVIRI sensor): detection indexes. Data observations from MTG (FCI sensor): detection indexes. <p>E-CONTRAIL consortium plan using 3 different algorithms to detection ice particles:</p> <ol style="list-style-type: none"> Brightness temperature difference (BTD) Covariance-Based Retrieval Algorithm (COBRA) Deep machine learning (DML) technique based on RGB ash composite
Historical aerial traffic	Timeline flight trajectory	<ul style="list-style-type: none"> ADS-B messages broadcasted basically by all airliners using their transponders. On the base of those messages, received by ground-based receivers, the trajectories of flight can be reconstructed. EUROCONTROL, as the Network Manager (NM), provides operational stakeholders with the most accurate picture of past and future pan-European air traffic demand from several years ahead until the day before operations. This includes both filed flight plans and flights based on radar tracks. All this information is managed by and can be accessed via the Research R&D archive, which is accessible.
Historical NWP data	Model outputs	<ul style="list-style-type: none"> The Historical NWP model outputs from The European Centre for Medium-Range Weather Forecasts (ECMWF), provide ERA5 (Reanalysis v5). Forecast data is available at the Copernicus Data Service. ERA5 is the latest and most advanced reanalysis dataset from ECMWF, covering the period from 1979 to the present with high spatial (approximately 30 kilometres) and temporal (hourly) resolution. ERA5 offers a comprehensive set of atmospheric, land, and oceanic variables at different altitudes and levels, making it suitable for a wide range of applications, including climate research, weather forecasting, and more. ERA5 provides numerous combinations of variables for the selection of input parameters for the Deep Learning models. The selection of optimal input variables enables the best possible outcomes in assessing the impacts of aviation-induced contrails over radiative forcing and climate change.

Table 3: E-CONTRAIL data input (description)

1.2.2 Output data

Four ways to exchange output data are considered within the E-CONTRAIL Project:

1. First, the consortium members will produce documentation, mainly related to data analysis, methodologies, processes and so on. The Data produced in that way will be presented as Open

Access under free licence whenever possible.³ All papers, reports and articles will be presented as Green-Open Access via UC3M's institutional repository *e-archivo* (refer to Section 3), which is linked to OpenAire. Thus, both OpenAire platform and the E-CONTRAIL Webpage will host this documentation. Furthermore, there will be some documents, that will be also published as Gold-Open Access. See Section 3 for the allocation of recourses to cover it.

2. Second, E-CONTRAIL will produce libraries/software as output of some of the activities of the project (namely, associated to WP1, WP2, and WP3). They will follow the software management plan (see [1]), including languages and standards specified therein. Some of these libraries might be offered to the community in open source (eventually with some restrictions), while others might be kept closed for further exploitation by the partners of E-CONTRAIL. The decision of whether to disclose this libraries/software or not will be taken by the partners upon discussion and deliberation within the General Assembly (see [2] for the management structure of E-CONTRAIL Project).
3. Third, the consortium will create forecast model outputs using a homogenised format (i.e., localisation of area with potential creation of ice crystal and AIC). The format chosen is the NetCDF (Network Common Data Form - <https://unidata.ucar.edu/netcdf>) format, which relies on a set of software libraries and self-describing, machine-independent data formats that support the creation, access, and sharing of array-oriented scientific data³. This open standard format is a project hosted by the Unidata program at the University Corporation for Atmospheric Research (UCAR). Chief source of NetCDF software, standards development, updates are freely available. NetCDF Classic and 64-bit Offset Format are an international standard of the Open Geospatial Consortium. An alert data product in NetCDF format is:
 - Self-Describing: a NetCDF file includes information about the data it contains.
 - Portable: a NetCDF file can be accessed by computers with different ways of storing integers, characters, and floating-point numbers.
 - Scalable: small subsets of large datasets in various formats may be accessed efficiently through NetCDF interfaces, even from remote servers.
 - Appendable: data may be appended to a properly structured NetCDF file without copying the dataset or redefining its structure.
 - Sharable: one writer and multiple readers may simultaneously access the same NetCDF file.
 - Archivable: access to all earlier forms of NetCDF data will be supported by current and future versions of the software.
4. Four, the consortium will make some development for the visualisation of E-CONTRAIL contrails and AIC detection, forecasts, and radiative forcing for identified flight trajectories.

³ Note that, as recognized by the Commission, participating in the ORD Pilot does not necessarily mean opening all your research data. Rather, the ORD pilot follows the principle "as open as possible, as closed as necessary" and focuses on encouraging sound data management as an essential part of research best practice. There might good reasons to keep some or even all research data generated in a project closed. In the case of sharing data under restrictions, explain why, clearly separating legal and contractual reasons from voluntary restrictions.

³ <https://www.unidata.ucar.edu/software/netcdf>

Table 4 summaries the four data outputs created as part of E-CONTRAIL project.

Category	Data	Description
Contrails and AIC	Historical detection	Data archive of contrails and AIC detections will be retrieved using a solution which combined the detection of ice crystal and historical aerial traffic data.
	Model Forecasts	Historical forecasts for regions favourable to the creation of contrails.
	Climate assessment	Historical climate impact (maps of radiative forcing at the location of high-altitude cirrus clouds) with respect to data archive of contrails detected and forecasts.
	Visualisation	For selected case studies, E-CONTRAIL dashboard visualises contrails detection, forecasts of region with contrails and climate impact of detected and forecasted contrails.

Table 4: E-CONTRAIL data output (description)

1.3 Origin, access, and size of the data

Category	Data	Origin	Access	Size
Particles in suspension (ice crystal)	Level-1 satellite observations from SEVIRI / MSG	EUMETSAT hubs	RMI has access to the Level-1 data and will calibrate them to physical values	~15 GB / day
	Level-2 satellite observations from SEVIRI / MSG (3 solutions of detection index)	UC3M	UC3M creates detection index using BTM	~5 GB / day
		BIRA	BIRA creates detection index using COBRA	~5 GB / day
		UC3M	UC3M creates detection index using DML	~5 GB / day
	Level-1 satellite observations from FCI / MTG	EUMETSAT hubs	RMI has access to the Level-1 data and will calibrate them to physical values	~220 GB / day
	Level-2 satellite observations from FCI / MTG (3 solutions of detection index)	UC3M	UC3M creates detection index using BTM	~70 GB / day
		BIRA	BIRA creates detection index using COBRA	~70 GB / day
		UC3M	UC3M creates detection index using DML	~70 GB / day

Historical aerial traffic	Timeline flight trajectory	ADSB	Uc3M has acquired a 2-years licence to access ADSB data from FlightRadar24	~10Gb/day
Historical NWP data	Model outputs (2 models)	ECMWF	RMI will provide a copy of historical forecasts.	~2 GB/day

Table 5: E-CONTRAIL data input (origin, access, and size)

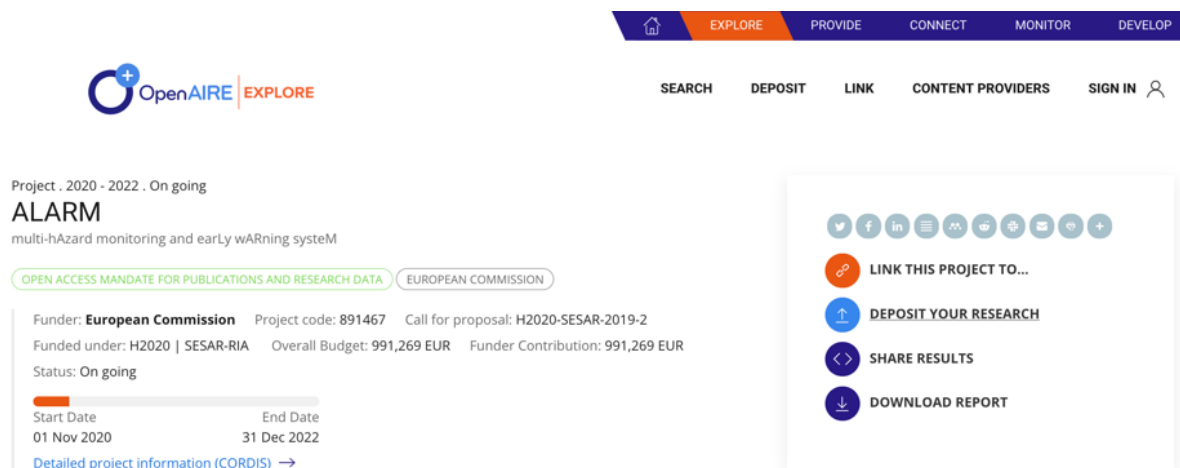
Level-1 data are quality-controlled and radiometrically calibrated, spectrally characterised, geometrically characterised, then annotated with satellite position and pointing, landmarks, and preliminary pixel classification.

Level-2 data consist of Earth located swath-based measurements from Floris, OLCI, and/or SLSTR at their native spatial resolution converted to geophysical parameter values using models and auxiliary datasets.

1.4 Re-use of data

Data re-use is a concept that involves using research data for a research activity or purpose other than that for which it was originally intended. Considering all the data input needed for E-CONTRAIL (see Table 5) will be re-used from its original purpose.

E-CONTRAIL will also foster its output data reuse by submitting data to OpenAire (see Figure 3 as an example) and by ensuring that data are described and comprehensible to any other researcher who may want to use it. When required, we will always follow metadata schemas to describing datasets so that they can be reused over time. Revision of the present document (see Section 1.3) will incorporate detailed information, including metadata, about the data used for E-CONTRAIL research activities. Similarly, software libraries will follow standards, will be documented, and will include test cases to facilitate its re-use.



Project . 2020 - 2022 . On going

ALARM

multi-hazard monitoring and early wARning systeM

OPEN ACCESS MANDATE FOR PUBLICATIONS AND RESEARCH DATA | EUROPEAN COMMISSION

Funder: **European Commission** Project code: 891467 Call for proposal: H2020-SESAR-2019-2
 Funded under: H2020 | SESAR-RIA Overall Budget: 991,269 EUR Funder Contribution: 991,269 EUR
 Status: On going

Start Date: 01 Nov 2020 End Date: 31 Dec 2022
[Detailed project information \(CORDIS\) →](#)

LINK THIS PROJECT TO...
 DEPOSIT YOUR RESEARCH
 SHARE RESULTS
 DOWNLOAD REPORT

Figure 3: Open Aire portal corresponding to a Project coordinated by UC3M (ALARM) as an example of the one to be created for E-CONTRAIL.

1.5 Data utility

Data utility is a measure of the business value attributed to data within specific usage contexts. Though utility is typically measured in monetary units, in the context of research, data utility can be measured in terms of knowledge re-use by different target communities.

The target audiences for E-CONTRAIL output data utility will be:

1. General public.
2. Undergraduate and graduate students.
3. Research and scientific community.
4. Stakeholders.
5. Institutions.

E-CONTRAIL will be making available different types of data to maximize the data utility for the different target groups. In particular:

For the General public, E-CONTRAIL will be realising a Concept Note, press-releases, and periodical updates with key results. The documentations, yet key data (e.g., videos, animations) will be released following the FAIR principles stated in Section 2. The goal is to ensure that the research activities are made known to society in such a way that they can be understood by non-specialists, thereby improving the public's understanding of science and technology. Establish a concern on how European collaboration and funding contributes to society.

For undergraduate and graduate students, E-CONTRAIL will participate in science weeks, researchers' nights, and incorporate some of the results in the teaching activities of the academic institutions, including Bachelor and Master Theses. Similarly, the documents and key data (e.g., videos, animations), including those used for teaching activities, will be released following the FAIR principles stated in Section 3. The goal is to wake up interest to follow scientific careers, especially to join E-CONTRAIL groups and follow PhD studies and develop BSc and MSc thesis.

For the research and scientific community, E-CONTRAIL main activity will be the publication of scientific papers and the development of open-source libraries of its re-use. The papers (including relevant metadata) will be published open source (as a general rule, following green-open access schemes at OpenAIRE; in some cases, following gold open-access schemes); some of the libraries will be also shared with an open-source license (including standardized documentation), e.g., by uploading them in a GitHub repository. The goal is maximizing the dissemination, including other scientific disciplines; enhance excellence and scientific reputation; find follow-up ideas and collaborations.

For the Stakeholders, besides scientific papers, and open-source libraries, which could be used on a free-basis to further develop commercial tools, a dashboard demonstration will be also developed during the project. These services will be mainly associated to the sharing of output products (shared as a NetCDF file). Fostering the commercial use of E-CONTRAIL output products/solutions is not in the objective of E-CONTRAIL.

For the Institutions, E-CONTRAIL intends to show (gathering all the data to be generated targeted to the previous groups) that the research and the re-usability of data is indeed interesting for the aviation community. The goal, in the end, is to draw the attention of different institutions, e.g., European Commission, SESAR, groups of interest, to make them aware of E-CONTRAIL results and make them visible to their agendas. This will facilitate the allocation of more funding and the revision/modification of standards.

2 FAIR data

FAIR data refers to a set of guiding principles to make data Findable, Accessible, Interoperable, and Reusable. The term was first presented at a Lorentz Workshop in 2014⁴ and the final principles were published in Nature in 2016⁵. The E-CONTRAIL Project will follow those principles as shown below.

2.1 Making data findable, including provisions for metadata

The findability of data is the ability to locate information by other users. To fulfil this principle, E-CONTRAIL will provide the necessary metadata to locate the information. As referred in the FAIR principles, the necessary metadata will include, wherever it is possible, the assignment of a unique and persisted identifier to each dataset, a data description with rich metadata, the inclusion of the identifier in the metadata and register on a searchable index the metadata.

2.1.1 e-archivo (institutional repository of UC3M)

Sharing data resources to different target audiences, including citizens, requires the provision of metadata. This is irrespective of the data being open or not. E-CONTRAIL will rely on *e-archivo* (<https://e-archivo.uc3m.es/>), the institutional repository of UC3M, to deposit in OpenAire.

e-archivo, the institutional repository of UC3M, has as one of its objectives the integration, conservation and preservation of the intellectual production resulting from the research and academic activity of the UC3M university community, in digital format.

The digital preservation of e-Archive contents guarantees the ability to use these contents in the future. The technical measures that are carried out are:

- Backups.
- Hardware and software updates.
- Maintenance of a mirror server.
- Format migration if required.

2.1.2 Unique and persistent identifier

Unique and persistent identifier: The items in e-File are associated with a unique and persistent identifier. The identification of the digital object guarantees access to the resource permanently. This means that this identifier will not be modified by server or domain changes on the Internet, etc.

The identification system used is "Handle" developed by the CNRI (Corporation for National Research Initiatives). The identifier consists of three parts that make up a unique and persisted identifier:

⁴ <https://www.force11.org/group/fairgroup/fairprinciples>

⁵ <https://www.nature.com/articles/sdata201618>



Figure 4: Unique and persisted identifier: Server (Servidor); Repository ID (Identificador del repositorio); Item ID (Identificador del ítem)

2.1.3 Document Format

The type of document format is closely related to the possibility of retrieving and accessing its contents in the future, due to technological obsolescence. e-File supports different types of commonly used standard media.

However, e-Archivo only guarantees the support of open standards that allow the retrieval and migration of document content. Figure 5 describes the supported formats:

MIME type estándar	Descripción	Extensiones
application/pdf	Adobe PDF	pdf
audio/x-aiff	AIFF	aiff, aif, aifc
audio/x-mpeg	MPEG Audio	mpa, abs, mpega
text/plain	Text	txt
text/html	HTML	htm, html
text/xml	XML	xml
text/richtext	Rich Text Format	rtf
text/csv	Comma Separated Values	csv
text/tab-separated-values	Tab Separated Values	tab
image/jpeg	JPEG	jpeg, jpg
image/gif	GIF	gif
image/png	PNG	png
image/tiff	TIFF	tiff, tif
video/mpeg	MPEG	mpeg, mpg, mpe

Figure 5: Type of archives

2.1.4 Metadata

E-Archive documents are described with metadata. Metadata is data associated with a digital document that collects mainly descriptive information (author, title, etc.), although it may include information on administration (creation of the resource, rights, access control ...) and preservation (type of format, etc.).

The metadata describes, identifies, and allows access to the resource. The metadata scheme used in e-Archive is Dublin Core Metadata Terms.⁶

innovation in metadata design, implementation & best practice

Join DCMII!

Dublin Core™ Metadata Initiative

DCMI Metadata Terms	
Title:	DCMI Metadata Terms
Creator:	DCMI Usage Board
Identifier:	http://dublincore.org/specifications/dublin-core/dcmi-terms/2020-01-20/
Date Issued:	2020-01-20
Latest Version:	https://www.dublincore.org/specifications/dublin-core/dcmi-terms/
Version History:	https://www.dublincore.org/specifications/dublin-core/dcmi-terms/release_history/
Document Status:	This is a DCMI Recommendation.
Description:	This document is an up-to-date specification of all metadata terms maintained by the Dublin Core Metadata Initiative, including properties, vocabulary encoding schemes, syntax encoding schemes, and classes.

Figure 6: Dublin Core Metadata Terms

The e-File metadata may be used under the conditions of the Creative Commons Attribution license (by): any collector or user may make use of our metadata with the only requirement that the source and identifier of the original record be mentioned in e -Archive.

The full text of the documents may be reused as specified in each of the items. In no case may it be collected by service providers. A link must be made to the unique and persisted identifier of each item in our repository.

2.2 Making data accessible

E-CONTRAIL, as previously stated, will produce four types of data outputs, namely: documentation (scientific and outreach), libraries/software, products (detection, forecasts and climate impact of contrails), and visualisation via demonstration dashboard. Data and results will be, whenever it is possible and in accordance with the Grant Agreement Art.26 [4], published openly and provide access.

2.2.1 Repository and protection of data

As for the decision on whether to published data or protect them, it should be stated that participating in the ORD Pilot does not necessarily mean opening all the research data, as recognized by the Commission. Rather, E-CONTRAIL will follow the ORD pilot principle "as open as possible, as closed as necessary", which is encouraging sound data management as an essential part of research best practice. The PMP establishes the procedures for internal review of documentation, including scientific publications (which might be subject to protection) and the eventual release of a library/software as

⁶ <https://www.dublincore.org/specifications/dublin-core/dcmi-terms/>

open source (conversely, protected for further exploitation). The governance structure of E-CONTRAIL (refer to [3]) includes an exploitation board and a dissemination board (both of them included in the General Assembly board). Any decision to be taken, must be previously discussed at these forum levels.

2.2.2 Data and accessibility

As for documents, there is no specific software requirement or documentation needed:

- Reports (in pdf format, including metadata) will be accessible mainly via the project webpage.
- Scientific papers will be made available via OpenAIRE (in pdf format, including metadata) and <https://arxiv.org/>, besides the website of the project. OpenAIRE has an arrangement with UC3M (the Project Coordinator) institutional repository, *e-archivo*.

As for open-source software libraries, they will be stored in GitHub, including the standardized documentation needed to use and access to the available data within it. The documentation will include software requirements (e.g., Python version and third-party libraries required) and instructions on how to install it and run available tests/examples.

Products developed during the E-CONTRAIL project (contrails and AIC detection, forecasts of contrails area and climate assessment) will be accessible in NetCDF format (via email exchange, personal communication, and project webpage). These products will be accessible on E-CONTRAIL dashboard demonstration website (address to be defined).

2.2.3 Metadata

Meta data will be accessible mainly via project webpage (in link data access/availability). Metadata is data associated with a data product outputs (contrails and AIC detection, forecasts of contrails area and climate assessment). It collects mainly descriptive information (title, type of product, time, geolocation, etc.), although it may include information on administration (name of provider, rights, access control ...) and preservation (type of format, etc.).

2.3 Making data interoperable

E-CONTRAIL will make its data interoperable, i.e., allowing data exchange and re-use between researchers, institutions, organisations, including different countries. For that purpose, we will use standards for the metadata for those documents in OpenAIRE, as indicated in Section 2.1.4. The libraries and software will be also developed following standards (see more details about software management in the [3]), e.g., we will follow the PEP8 standard in Python developments, including coding, documentation, and test standards to ensure interoperability.

2.4 Increase data re-use

To increase data re-use, E-CONTRAIL will first follow a decision process on whether to protect certain data or make them freely available. This has been covered in Section 2.2.

In the case of open release, we will add an open license.

An open license is a notice that grants the recipient of a creation (can be a software, a document, or other creations) extensive rights (with certain conditions) to use, modify and redistribute that creation. These actions are usually prohibited by copyright law, but the rights-holder (usually the author) of a creation can remove these restrictions by accompanying the creation with a license which grants the recipient these rights. Using such a license is free as conferred by the copyright holder. However, the way authorship must attribute, the way of further sharing, licensing derived products, and eventually commercialized is included in the open license.

The permissions and limitations of a free license are classified according to the following subjects:

- Linking - linking of the licensed code with code licensed under a different licence (e.g., when the code is provided as a library)
- Distribution: distribution of the code to third parties
- Modification: modification of the code by a licensee
- Patent grant: protection of licensees from patent claims made by code contributors regarding their contribution, and protection of contributors from patent claims made by licensees.
- Private use: whether modification to the code must be shared with the community or may be used privately (e.g., internal use by a corporation)
- Sublicensing: whether modified code may be licensed under a different licence (for example a copyright) or must retain the same licence under which it was provided
- TM grant: use of trademarks associated with the licensed code or its contributors by a licensee.

Recent tools like the European Commissions' Join up Licensing Assistant⁷, makes possible the licenses selection and comparison based on more than 40 subjects or categories, with access to their Software Package Data eXchange (SPDX) identifier and full text.

Though E-CONTRAIL will be exploring the most adequate license for each case, tentatively, E-CONTRAIL:

- For documents, will rely on Creative Commons Attribution 4.0 Int. (see Fig. Figure 7)
- For software libraries, will rely on GNU General Public License v3.0 (see Figure 8)

⁷ <https://joinup.ec.europa.eu/solution/joinup-licensing-assistant/joinup-licensing-assistant-jla>

CC-BY-4.0 Creative Commons Attribution 4.0 International (CC-BY-4.0)

Can	Use/reproduce, Distribute, Modify/merge, Sublicense, Commercial use
Must	Incl. Copyright, State changes
Compatible	None N/A, Permissive, Multilingual, For data
Law	Not fixed/local
Support	Strong Community

Licence comment:

The less restrictive Creative Commons Attribution licence is this international version 4, which gives recipients maximum freedom to do what they want with the work of the licensor. Recipients redistributing the work must give credit to the original author of the work (= attribution) and state changes if any, including a URL or link to the original work, this CC-BY licence and a copyright notice. Author can request to remove any attribution given information. Recipients re-distributing the work to third parties may not apply legal terms or technological measures (like Tivoisation) that legally restrict the rights granted by the licence. OKF (Open Knowledge Foundation) recommends this licence...

Figure 7: Creative Commons Attribution 4.0 International

GPL-3.0-only GNU General Public License v3.0

Can	Use/reproduce, Distribute, Modify/merge, Commercial use, Use patents, Place warranty
Must	Incl. Copyright, State changes, Disclose source, Copyleft/Share a., Include licence
Cannot	Hold liable, Use trademark
Compatible	For software
Law	Not fixed/local
Support	Strong Community, OSI approved, FSF Free/Libre

Licence comment:

Fundamental revision of the GPL produced in 2007. OSI states GPL (without indicating version number) as supported by a strong community (the Free Software Foundation - FSF). You may copy, distribute and modify the software as long as you track changes/dates in source files. Any derivative including (i.e. via compiler) GPL-licensed code must also be made available under the GPL-3.0 along with build & install instructions. Lack of compatibility is a known GPL-3.0 issue, making its use problematic when a project combines multiple sources and stakeholders: at the exception of AGPL-3.0, the licence is compatible with no other (even the GPL-2.0) and, according to the FSF, linking other software to GPLed works extends the GPL coverage to these other programs.

Figure 8: GNU General Public License v3.0

3 Allocation of resources

As defined in the SESAR JU Horizon grant agreement [3], costs related to open access to research data are eligible for reimbursement.

Accordingly, and in order to make E-CONTRAIL data FAIR, the project has allocated:

- 2000€ per partner to gold open access publications.

Moreover, the following activities are also covered:

- The publication of research articles in Green Access (via OpenAIRE) is covered by institutional resources. For instance, UC3M's institutional repository *e-archivo* is directly linked to OpenAIRE.
- The publications of open-source libraries in GitHub is free, including its maintenance.
- UC3M team has a Server with capacity over 100TB for data storage where the research data needed for the project will be allocated.
- BIRA has also infrastructure to store the data needed for this project.

UC3M, as coordinator of the project, in close cooperation with BIRA, as host of the service platform, will be responsible for the management of data.

Preserved datasets will need to be updated and this means a data preservation policy and process will need to be defined. A central consideration for any long-term DMP is the cost of preserving that data and what will happen after the completion of the project. Preservation costs may be considerable depending on the exploitation of the project after its finalization. Examples include:

- Personnel time for data preparation, management, documentation, and preservation, Hardware and/or software needed for data management, backing up, security, documentation, and preservation,
- Costs associated with submitting the data to an archive,
- Costs of maintaining the physical backup copies (disks age and need to be replaced).

These topics will be discussed during the Project at the General Assembly level. If any relevant conclusions or recommendations are reached, they will be included in potential updates of this DMP. This will be updated in future version of the present document.

4 Data security

4.1 Data sharing

Several datasets will be used during E-CONTRAIL, which will share those datasets with respect to the legal and Intellectual Property Rights (IPR) constraints. Specifically, we have to consider where, how, and to whom the data could be made available. This information will be updated in the updates of this document, together with the description of the data.

Datasets will be shared either by opening a specific API for third parties to fetch datasets from the E-CONTRAIL store, or via making such datasets downloadable as single files. In any case, the methods used to share data will be dependent on several factors such as the type, size, complexity, and sensitivity of data.

The following general conditions for accessing and using the data apply:

- The Dataset provided will never be the raw data obtained from operational sources, but the result of the fusion and processing performed the E-CONTRAIL partners providing the data.
- The Dataset can only be used to achieve the research purposes stated in the E-CONTRAIL Technical Description of the Action. No other use is allowed.
- The user is not allowed to create or derive new datasets from the original one.
- The Dataset will be stored physically in the premises of UC3M and/or BIRA, as host of E-CONTRAIL's dashboard. No physical copy of the Dataset will be provided. The users will not store, copy, or otherwise move the Dataset (physically or logically) to databases or systems outside the premises of UC3M/BIRA.
- Access to the Dataset will be granted as needed by the research members of the consortium either for implementing their own tasks under the E-CONTRAIL action or for exploiting their own results.
- Access to the Dataset will be granted under a specific confidentiality agreement that will clearly identify the Dataset users, purpose, usage timeframe and any specific clauses that might be needed.
- The Dataset access confidentiality agreement will be signed before the E-CONTRAIL of the work to set the conditions of this access. UC3M/BIRA (or, eventually, other E-CONTRAIL partners providing the data) is committed to provide access in a secure way, royalty-free, to achieve the research goals of the project in an efficient way.

4.2 Data storage and preservation

E-archivo aims to be the digital memory of the scientific and academic production of UC3M. The official adhesion of the University to the open access movement (Open Access), by signing the Berlin Declaration - <https://openaccess.mpg.de/Berlin-Declaration> - in June 2006, commits the repository to ensure this objective.

Authors may request the correction or modification of the metadata, if there is an error, by contacting the e-Archive team by email or telephone. They may also request the deposit of a new version of the full text of a work.

e-Archivo will keep the documents that the researchers have authorized to deposit, even after leaving the University.

5 Ethics

Ethical issues have been identified on personal data concerning the participants in the Advisory Board and Stakeholder involvement. The treatment of the associated data will be addressed in E-CONTRAIL's Deliverables D3.1 [4] and D5.2 [2].

6 Other issues

1. Commission EU, *Horizon Europe: Data Management Plan Template v1.0* 05.05.2021.
2. Soler, M. et al., D5.1 PMP – Project Management Plan. *E-CONTRAIL Project (GA- 101114795). D5.1.* 2023.
3. Commission EU, *CA – Consortium Agreement E-CONTRAIL project. Version 1.0, 8 Mars 2023.* E-CONTRAIL project (GA- 101114795).
4. Otero, M. et al., D3.1 H – Relevant data for deep-learning models. Report summarizing the relevant data sources and their usage in the various models. *E-CONTRAIL Project (GA- 101114795). D3.1.* 2023.

